De la difficulté de garder ses amis (quand on a des ennemis) !

Guillaume Ducoffe, Dorian Mazauric, Augustin Chaintreau

▶ To cite this version:

Guillaume Ducoffe, Dorian Mazauric, Augustin Chaintreau. De la difficulté de garder ses amis (quand on a des ennemis) !. 15èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel), May 2013, Pornic, France. pp.1-4. hal-00815680

HAL Id: hal-00815680

https://hal.science/hal-00815680

Submitted on 25 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

De la difficulté de garder ses amis (quand on a des ennemis) !

Augustin Chaintreau[1] and Guillaume Ducoffe[1,2] and Dorian Mazauric[1,3]

[1] *Columbia University, Computer Science Department, New York*
[2] *École Normale Supérieure de Cachan*
[3] *Laboratoire d'Informatique Fondamentale de Marseille, équipe ACRO*

Nous nous intéressons aux processus locaux de formation des groupes de partage d'information dans les réseaux sociaux. Le réseau est modélisé par un graphe arête-valué où le poids (positif ou négatif) d'une arête représente l'utilité que les deux sommets ont s'ils se trouvent dans un même groupe. Nous supposons que les groupes de partage forment une partition des sommets. Le processus local de formation des groupes est basé sur l'optimisation de l'utilité individuelle de chaque sommet. Soit k la taille maximale d'une coalition. Un sous-ensemble d'au plus k personnes peut rejoindre un groupe existant ou créer un nouveau groupe si, et seulement si, leurs utilités respectives augmentent strictement. Ce changement est appelé une k-déviation. Une partition est k-stable si, et seulement si, il n'y a pas de k-déviation possible.

Kleinberg et Ligett ont montré que si les poids sont $-\infty$ et 1 (ennemis et amis), alors il existe toujours une partition k-stable pour tout $k \geq 1$. Ils ont également montré que pour $k \in \{1,2,3\}$, le nombre maximum de déviations avant d'atteindre une partition k-stable est polynomial. La polynomialité du temps de convergence dans le cas le pire pour $k = 4$ a été laissée comme problème ouvert. Nous montrons que ce dernier peut être $\Omega(n^{c\log(n)})$ avec c une constante et n le nombre de sommets. De plus, nous prouvons une formule close pour $k \in \{1,2\}$, une meilleure borne inférieure pour $k = 3$ et des résultats d'existence et de complexité pour des poids généraux.

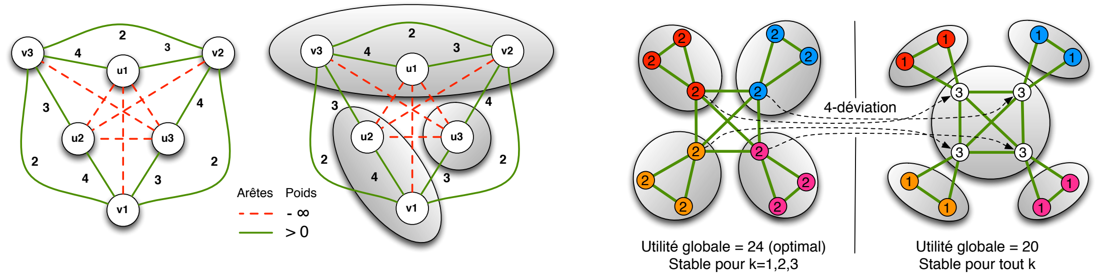
Keywords: partage d'information, réseaux sociaux, algorithme local, stabilité, temps de convergence.

1 Introduction.

Contexte et motivation. *"Qui reçoit de moi une idée augmente son instruction sans diminuer la mienne."* De cet idéal énoncé par Thomas Jefferson on peut retenir que partager de l'information produit parfois un bénéfice mutuel. L'importance d'être *bien* informé motive une palette d'activité sociale (groupes, communautés, réseautage), récemment transformés par des outils en lignes, avec un net risque de surexposition. Facebook, Twitter, Weibo simplifient le partage et suppriment les frontières géographiques ou de classes. Mais ils transforment aussi le contexte de cette information, traditionnellement contrôlée par la diffusion entre cercles et groupes sociaux qui tiennent à l'écart ceux que l'on ne veut pas tenir informés de tout, voire du tout.

Comment se constituent – ou pourrait mieux se constituer – des groupes de partage d'information dans un graphe où les arêtes représentent soit un bénéfice mutuel, soit une incompatibilité ? Éviter les paires incompatibles en formant des partitions où l'information se propage peut sembler efficace. Assigner ces groupes par un algorithme de coloration de graphe, par contre, semble *simpliste* et peut se révéler *instable* car les noeuds, suivant leurs intérêts, peuvent dévier vers une configuration qui les favorisent. Nous étudions la *dynamique*, encore inconnue, de ce jeu distribué.

Modélisation du réseau social et des groupes. Le réseau est modélisé par un graphe arête-valué $G = (V,E,w)$ avec V représentant l'ensemble des utilisateurs. Le poids $w_{u,v}$ (positif ou négatif) entre deux sommets u et v représente l'utilité engendrée pour u et v si ces deux derniers partagent de l'information. Nous supposons que le graphe est complet en ajoutant des arêtes de poids 0 si nécessaire. Nous notons $\mathcal{W}$ l'ensemble des poids pris par les arêtes. La Figure 1(a)



(a) L'instance de gauche représente un graphe arête-valué avec $\mathcal{W} = \{-\infty, 2, 3, 4\}$. La partition décrite à droite est 1-stable mais il n'existe pas de partition 2-stable pour cette instance.

(b) Exemple de 4-déviation avec $\mathcal{W} = \{-\infty, 1\}$. La partition de gauche est 3-stable mais pas 4-stable, et l'utilité globale est maximale. La partition de droite est 4-stable mais l'utilité globale a diminué.

FIGURE 1: Deux instances du problème de partage d'information dans les réseaux sociaux.

(gauche) représente un exemple de graphe valué avec $\mathcal{W} = \{-\infty, 2, 3, 4\}$. Un poids $-\infty$ entre deux sommets représente le cas où les deux utilisateurs sont *ennemis* et ne veulent en aucun cas partager de l'information. Nous supposons enfin que les groupes de partage forment une *partition des sommets*. Autrement dit, un sommet appartient à un unique groupe. Étant donnée une partition C, l'utilité $f_u(C)$ d'un sommet u est la somme des poids des arêtes adjacentes à u dans son groupe. Formellement $f_u(C) = \sum_{v \in C(u)} w_{u,v}$ où $C(u)$ représente l'ensemble des sommets dans le même groupe que u. L'utilité du sommet v_3 pour la partition décrite dans la Figure 1(a) (droite) est 6. Notons que deux sommets ennemis (poids $-\infty$) ne sont jamais dans le même groupe mais que deux sommets liés par un poids négatif (autre que $-\infty$) peuvent être dans un même groupe en raison de compensations.

Processus local et optimisation individuelle. La dynamique du système est la suivante. Initialement, chaque utilisateur forme un groupe seul. Soit $k \geq 1$ la taille maximale constante d'une coalition. Étant donnés des groupes de partage, un sous-ensemble d'au plus k utilisateurs peut rejoindre un groupe existant ou créer un nouveau groupe si, et seulement si, leurs utilités respectives augmentent strictement. Ce changement est appelé une *k-déviation*. La partition est *k-stable* si, et seulement si, il n'y a pas de k-déviation possible.

La partition décrite dans la Figure 1(a) (droite) est 1-stable. En effet, aucun sommet ne peut créer un nouveau groupe seul et augmenter strictement son utilité. De plus, aucun sommet ne peut rejoindre (seul) un groupe existant car soit un ennemi y est présent soit il aurait une utilité strictement positive mais inférieure à son utilité courante. En revanche, cette partition n'est pas 2-stable car les deux sommets v_1 et v_2 peuvent former une coalition et rejoindre le groupe formé par u_3. L'utilité de v_1 passe de 4 à 5 et l'utilité de v_2 passe de 5 à 6. Après une telle 2-déviation, la partition obtenue est isomorphe à la précédente. Donc cette partition n'est pas 2-stable et, plus généralement, il est possible d'observer que cette instance n'admet pas de partition 2-stable.

Stabilité et temps de convergence. Un problème important est de caractériser les classes d'instances admettant une partition k-stable. Pour ces instances, il est alors intéressant de déterminer si une partition k-stable est atteignable par le processus dynamique opérant dans le système. Pour ces instances, nous nous sommes intéressés au temps de convergence du processus dynamique (nombre de k-déviations avant d'obtenir la k-stabilité) dans le pire des cas notamment.

Travaux existants. Kleinberg et Ligett ont montré dans [KL10] que si $\mathcal{W} = \{-\infty, 1\}$, alors il existe toujours une partition k-stable pour tout $k \geq 1$. Cette classe d'instances peut correspondre au cas où il n'y a que des *amis* et des *ennemis*, et où, pour des raisons de vie privée, un individu ne veut surtout pas partager d'information avec un ennemi. Pour ces instances particulières, ils ont montré que le temps de convergence est polynomial pour $k \in \{1, 2, 3\}$. Ils ont en revanche laissé ouvert le problème de la polynomialité du temps de convergence pour $k = 4$.

Contributions. Nous avons montré que, pour $k = 4$ et $\mathcal{W} = \{-\infty, 1\}$, le temps de convergence pouvait être $\Omega(n^{c\log(n)})$ avec c une constante et n le nombre d'utilisateurs. Pour $k \in \{1, 2\}$, nous avons prouvé une formule close pour le nombre exact de k-déviations dans le pire des cas (Tableau 1). De plus, nous avons montré que pour $\mathcal{W} = \{-\infty, 0, 1\}$, il existe pour tout réseau, une partition 1 et 2-stable (avec des temps de convergence polynomiaux) mais que certains réseaux n'admettent pas de partition 3-stable. Cette classe d'instances modélise le cas où les utilisateurs peuvent avoir des relations *neutres*. Plus généralement, nous montrons des résultats d'existence pour des poids généraux (Tableau 2) et nous prouvons que le problème de décider s'il existe une partition k-stable est NP-complet en général. L'intégralité de nos résultats et de nos preuves se trouve dans [DMC12].

2 Temps de convergence pour l'ensemble de poids $\mathcal{W} = \{-\infty, 1\}$.

Pour l'ensemble de poids $\mathcal{W} = \{-\infty, 1\}$, nous représentons toute partition C par un *vecteur de partition* $\Lambda = (\lambda_n, \ldots, \lambda_1)$ de taille n où λ_i représente le nombre de groupes de taille exactement i pour tout $1 \leq i \leq n$. Nous avons montré que, pour tout $k \geq 1$ et pour toute partition, n'importe quelle k-déviation augmente le vecteur de partition selon l'ordre lexicographique. Donc le processus local converge toujours vers une partition k-stable. Autrement dit, il n'y a pas de cycle de k-déviations pouvant empêcher le processus d'atteindre une partition k-stable.

Un paramètre important est alors le nombre maximal de k-déviations avant d'atteindre la k-stabilité. Étant donnés $k \geq 1$ et $n \geq 1$, nous notons $L(k,n)$ ce temps de convergence, dans le pire des cas, pour un graphe avec au plus n sommets. Notons que $L(k',n) \geq L(k,n)$ pour tout $k' \geq k \geq 1$ et pour tout $n \geq 1$.

Kleinberg et Ligett ont montré dans [KL10] que $L(1,n) = O(n^2)$, $L(2,n) = O(n^2)$ et $L(3,n) = O(n^3)$. Pour $k \in \{1,2\}$, leur preuve s'appuie sur le fait que la somme des utilités augmente strictement après n'importe quelle k-déviation. Le résultat s'obtient ensuite en remarquant que cette somme est bornée supérieurement par $O(n^2)$ car l'utilité individuelle est bornée supérieurement par $n-1$. Pour $k = 3$, ils ont montré que la somme des carrés des utilités augmente toujours strictement après n'importe quelle k-déviation (l'utilité globale peut ne pas augmenter) et donc que $L(3,n) = O(n^3)$. Ils ont en revanche laissé ouvert le problème de la polynomialité pour $k = 4$. Dans ce cas, toute fonction (additive) potentielle bornée par un polynôme peut décroître strictement pour certaines 4-déviations. Nous avons résolu le problème ouvert de [KL10] en montrant que $L(k,n)$ n'est pas polynomial pour tout $k \geq 4$.

Théorème 1 $L(4,n) = \Omega(n^{c \log(n)})$ *avec c une constante.*

En utilisant les techniques développées dans la preuve du Théorème 1, nous avons également prouvé une borne inférieure pour $k = 3$: $L(3,n) = \Omega(n^2)$. Nous avons également prouvé que $L(k,n) = O(e^{\sqrt{n}})$ améliorant la meilleure borne supérieure exponentielle connue. La preuve s'appuie sur le fait que le nombre de vecteurs de partitions différents est $O(e^{\sqrt{n}})$. Rappelons qu'un vecteur de partition augmente strictement selon l'ordre lexicographique après n'importe quelle k-déviation.

Pour améliorer les bornes supérieures, nous avons montré que nous pouvons nous ramener au cas $\mathcal{W} = \{1\}$. Autrement dit, pour tout $k \geq 1$ et pour tout $n \geq 1$, $L(k,n)$ est atteint pour $\mathcal{W} = \{1\}$. Nous avons alors prouvé que $L(1,n) = L(2,n)$ en montrant que cette valeur était égale à la longueur de la plus longue chaine dans le treillis des partitions. Utilisant les résultats de [GK86, GK93], nous déduisons la formule close :

Théorème 2 $L(1,n) = L(2,n) = 2\binom{m+1}{3} + mr$, *où r et m sont les uniques solutions de* $n = \frac{m(m+1)}{2} + r$, $0 \leq r \leq m$. *Cela implique que* $L(1,n) = L(2,n) \sim \frac{2}{3} n \sqrt{n}$ *quand n est grand.*

Le Tableau 1 résume les résultats des travaux existants et nos contributions.

3 Existence de partition k-stable pour des poids généraux.

Étant donné un ensemble de poids $\mathcal{W}$, $k(\mathcal{W})$ est défini de la manière suivante : pour tout $k \leq k(\mathcal{W})$, il existe une partition k-stable pour tout graphe et il existe un graphe qui n'est pas

k	Littérature	Nos résultats
1	$O(n^2)$	$\sim \frac{2}{3}n^{3/2}$
2	$O(n^2)$	$\sim \frac{2}{3}n^{3/2}$
3	$O(n^3)$	$\Omega(n^2)$
≥ 4	$O(2^n)$	$\Omega(n^{c\log(n)}), O(e^{\sqrt{n}})$

TABLE 1: Temps de convergence maximal $L(k,n)$ pour $\mathcal{W} = \{-\infty, 1\}$.

$\mathcal{W}$	$k(\mathcal{W})$
$\{-\infty, a, b\}$, $0 < a < b$	1
$\{-\infty, -\mathbb{N}, 0, 1\}$	2
$\{-\infty, b\}$, $b > 0$	∞
$\mathcal{W} \subseteq \mathbb{N}$; $\mathcal{W} \subseteq -\mathbb{N}$	∞
$-\mathbb{N} \cup \{N\}$ †	∞

TABLE 2: Existence de partition k-stable représenté par $k(\mathcal{W})$.

$(k(\mathcal{W})+1)$-stable. S'il existe une partition k-stable pour tout $k \geq 1$, nous définissons $k(\mathcal{W}) = \infty$ (*e.g.* $k(\{-\infty, 1\}) = \infty$).

Nous avons prouvé que $k(\mathcal{W}) \geq 1$ pour tout $\mathcal{W}$, c'est-à-dire qu'il existe toujours une partition 1-stable. Nous avons ensuite caractérisé les ensembles $\mathcal{W}$ tels que $k(\mathcal{W}) = \infty$. Précisément $k(\mathcal{W}) = \infty \Leftrightarrow \mathcal{W} = \{-\infty, b\}$ avec $b > 0$, $\mathcal{W} \subseteq \mathbb{N}$ ou $\mathcal{W} \subseteq -\mathbb{N} \cup \{N\}$ †. Nous avons également montré que si nous ajoutons des relations *neutres* entre individus à la classe d'instances ennemis et amis, alors la 3-stabilité n'est plus garantie. Autrement dit, pour $\mathcal{W} = \{-\infty, 0, 1\}$, il existe un graphe G qui n'admet pas de partition 3-stable et il existe toujours une partition 2-stable (avec temps de convergence polynomial). De plus, nous avons prouvé que $k(\{-\infty, 0, 1\}) = k(\{-\infty, -\mathbb{N}, 0, 1\})$. Enfin pour $\mathcal{W} = \{-\infty, a, b\}$, $0 < a < b$, nous avons montré que $k(\mathcal{W}) = 1$. Le Tableau 2 résume nos résultats.

Pour conclure, nous avons démontré que le problème de décider s'il existe une partition k-stable est NP-complet. Nous avons utilisé le problème de l'ensemble indépendant de cardinalité maximale dans notre réduction.

Théorème 3 *Pour tout $\mathcal{W}$ contenant $-\infty$ et pour tout $k > k(\mathcal{W})$, étant donné un graphe G avec les poids $\mathcal{W}$, le problème de décider s'il existe une partition k-stable pour G est NP-complet.*

4 Extensions et travaux futurs.

En plus de poursuivre notre étude pour des poids généraux, nous proposons d'étudier la formation des groupes de partage dans les réseaux sociaux lorsque les sommets peuvent appartenir à plusieurs groupes différents. Certains réseaux n'admettant pas de configuration k-stable dans le cas de la partition, peuvent maintenant avoir une configuration k-stable. Mais nos premiers résultats montrent également que cela peut rendre le réseau instable. En effet, la 3-stabilité n'est plus garantie pour l'ensemble de poids $\mathcal{W} = \{-M, 1\}$ lorsque les sommets appartiennent à deux groupes. Une autre extension intéressante est de prendre en compte des *utilités transitives*. Un utilisateur peut ne pas vouloir être dans un groupe comprenant simultanément une personne de son cercle amical et une personne de son cercle professionnel (alors que les utilités respectives sont intrinsèquement positives). Nous envisageons une modélisation à base d'hypergraphes.

Références

[DMC12] Guillaume Ducoffe, Dorian Mazauric, and Augustin Chaintreau. Convergence of coloring games with collusions. Technical report, Columbia University, August 2012. available at `www.cs.columbia.edu/~augustin/pub/DMC.TR13.pdf`.

[GK86] C Greene and D J Kleitman. Longest chains in the lattice of integer partitions ordered by majorization. *Eur. J. Comb.*, 7(1) :1–10, January 1986.

[GK93] Eric Goles and Marcos A. Kiwi. Games on line graphs and sand piles. *Theoretical Computer Science*, 115(2) :321 – 349, 1993.

[KL10] Jon M Kleinberg and Katrina Ligett. Information-Sharing and Privacy in Social Networks. *paper in progress (available at `arxiv.org/abs/1003.0469`)*, 2010.

†. Le poids $N > 0$ est plus grand que n fois la valeur absolue du poids négatif le plus petit (différent de $-\infty$). Nous pouvons considérer qu'il vaut $+\infty$ quand on le compare aux autres poids (de valeur finie).